

# Die relative Verteilung als Konzept zur Analyse von Gruppenunterschieden

Ben Jann, ETH Zürich

4. Juli 2007

Forschungskolloquium Empirische Sozialforschung  
Uni Konstanz

## Inhalt

- Einleitung: Analyse von Gruppenunterschieden
- Die relative Verteilung
- Empirische Beispiele

# Einleitung

- Die Analyse von Gruppenunterschieden bezüglich eines kontinuierlichen Merkmals, erfolgt meistens nur aufgrund einiger weniger, als zentral angesehener Masszahlen (i.d.R. Erwartungswert).
- Beispiel: Einkommen und Geschlecht. Wie wird das analysiert?
  - Mittelwertsdifferenz/Mediandifferenz
  - Kontrolle von Drittvariablen I: Geschlecht als Dummy-Variable in einem Regressionsmodell => konditionale Mittelwertsdifferenz
  - Kontrolle von Drittvariablen II (kontrafaktischer Ansatz): Dekomposition der (log.) Einkommensunterschiede in einen „erklärten Teil“ (Effekt der X-Unterschiede) und einen „unerklärten Teil“ („Diskriminierung“; Effekt der Unterschiede in den Koeffizienten) (Blinder 1973; Oaxaca 1973; etc.)

# Einleitung

- Solche Analysen sind zwar informativ, lassen aber auch viel im Dunkeln.
- Ansätze zu einer umfangreicheren Analyse der Gruppenunterschiede:
  - Semi-parametrische Erweiterung der Oaxaca-Dekomposition auf beliebige Masszahlen (Quantile, Streuung, etc.) mit Hilfe der Invertierung von Residuenverteilungen (Juhn, Murphy, Pierce 1993; Blau & Kahn 1996a)
  - Mit einem ähnlichen Ansatz: Analyse der Veränderung von Gruppenunterschieden unter Berücksichtigung der „allgemeinen“ Veränderung der Verteilung (Juhn, Murphy, Pierce 1991; Blau & Kahn 1992, 1996b, 1997)
  - Untersuchung der ganzen Verteilung mit Hilfe von Quantils-Regressionen (Buchinsky 1998); Erweiterung der Oaxaca-Dekomposition für Quantile (Machado & Mata 2005); nicht-parametrische Oaxaca-Dekomp. mit Hilfe von Matching (Nopo 2004)
  - Direkte Analyse von Differenzen in Dichtefunktionen; kontrafaktische Analyse mit Hilfe von Gewichten (DiNardo, Fortin, Lemieux 1996)

# Die relative Verteilung

- Weiterer (nicht-parametrischer) Ansatz zur Analyse der Unterschiede/Veränderung gesamter Verteilungen
- Literatur: Morris, Bernhardt, Handcock (1994), Bernhardt, Morris, Handcock (1995), Handcock & Morris (1998, 1999)
- Grundlegender Gedanke: Interpretation der Werte von Gruppe A als relative Positionen in der Verteilung von Gruppe B
- Ein Vorteil des Ansatzes ist, dass die Resultate weitgehend unabhängig sind von monotonen Transformationen des Merkmals.

## Relative Daten: Definition

- Seien  $F_0(y)$  und  $F(y)$  die Verteilungsfunktionen in der Referenzgruppe und der Vergleichsgruppe
- Die relativen Daten sind dann definiert als

$$R = F_0(Y)$$

(*grade transformation*);  $R$  in  $[0,1]$

- Verteilungsfunktion (CDF) von  $R$  ( $F^{-1}$  ist die Quantils-Funktion):

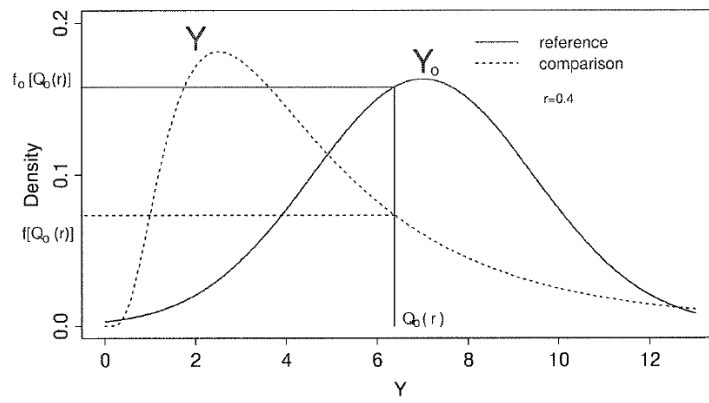
$$G(r) = F(F_0^{-1}(r)) \quad 0 \leq r \leq 1$$

- Dichtefunktion (PDF) von  $R$ :

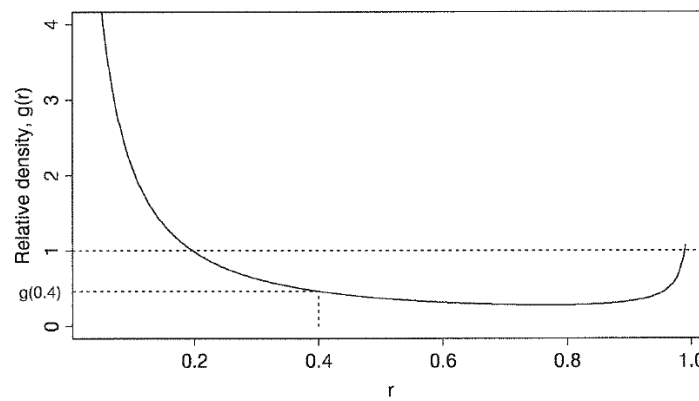
$$g(r) = \frac{f(F_0^{-1}(r))}{f_0(F_0^{-1}(r))} \quad 0 \leq r \leq 1$$

(entspricht dem Verhältnis der Dichten in den beiden Gruppen)

(a) PDF Overlay



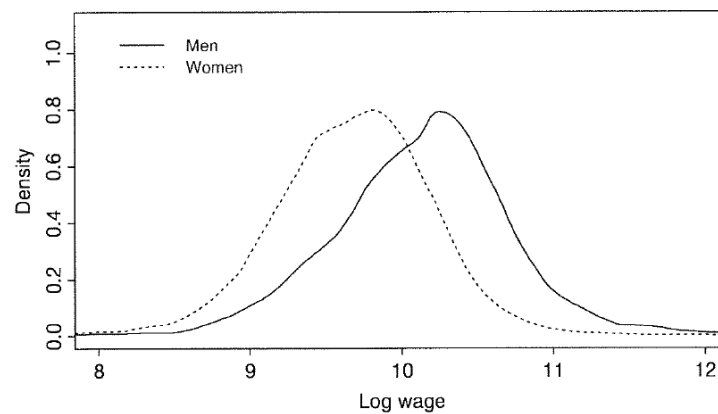
(b) Relative PDF



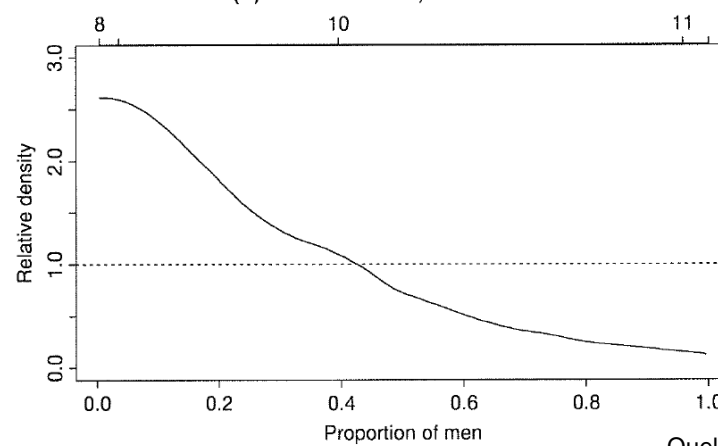
proportion of reference group

Quelle: Handcock & Morris (1999:23)

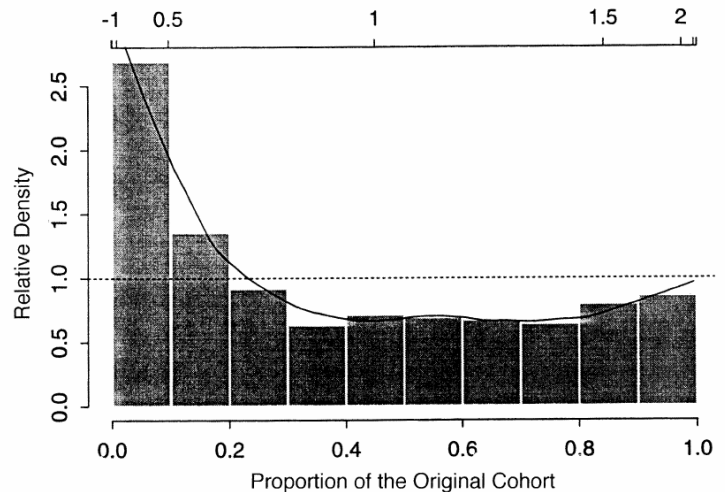
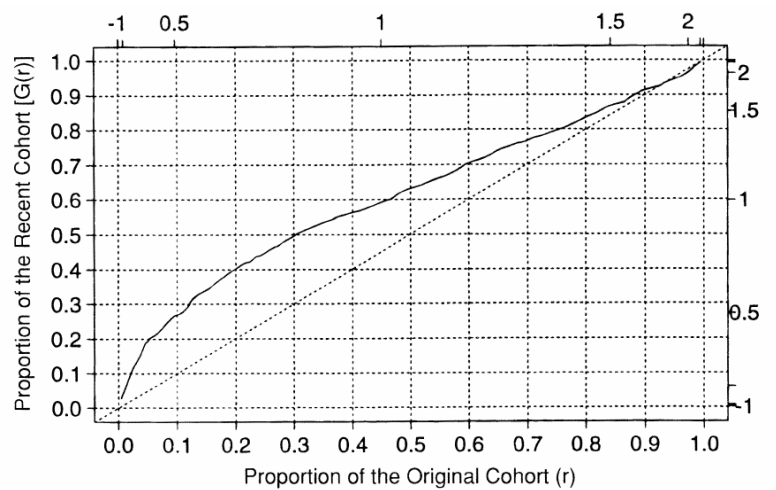
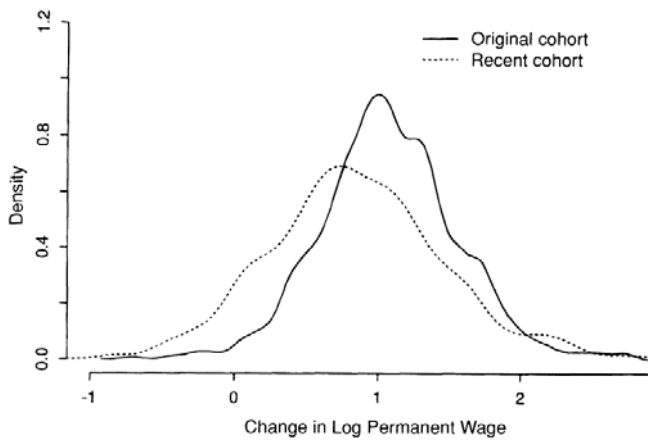
(a) Log Earnings PDFs for Men and Women, 1987



(b) Relative PDF, Women:Men



Quelle: Handcock & Morris (1999:24)



Quelle: Handcock & Morris (1998:58/65)

## Dekomposition der Unterschiede: zentrale Tendenz und Form

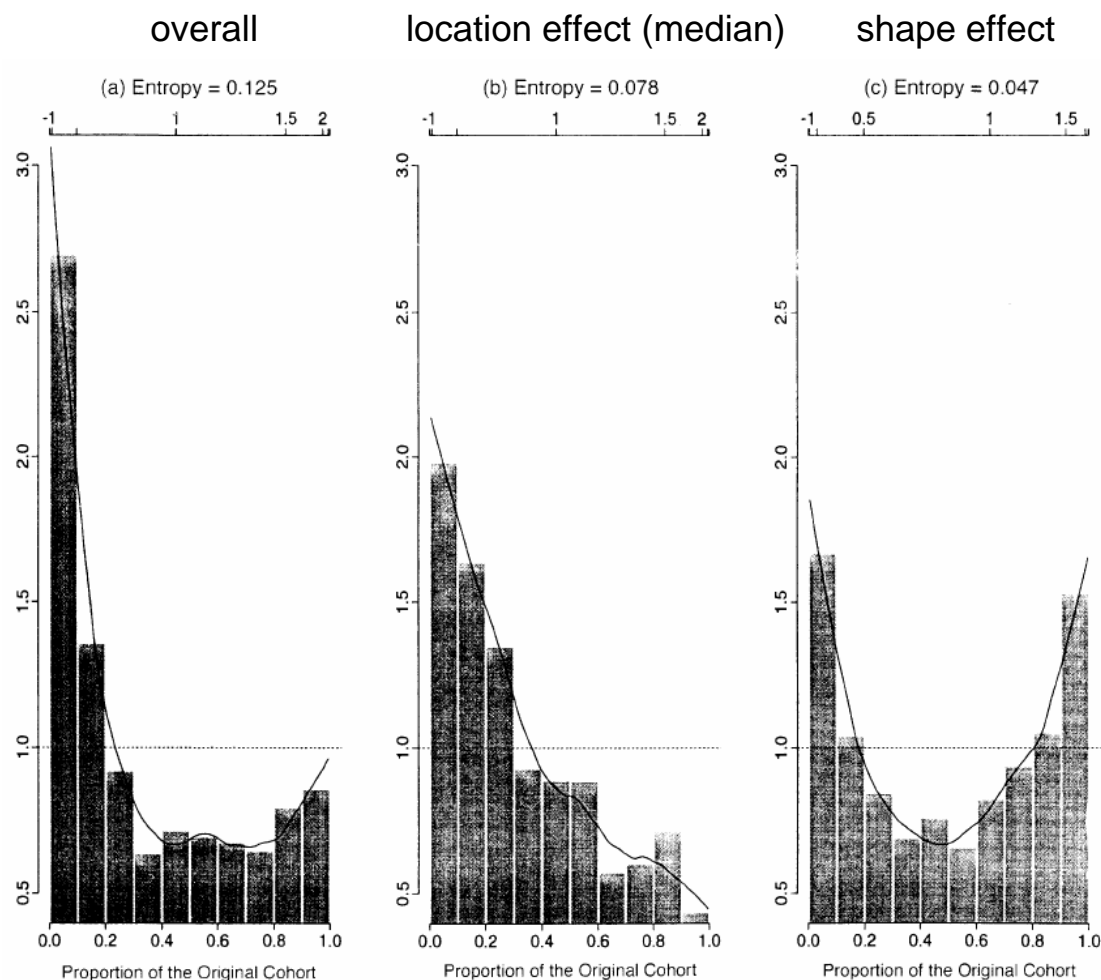
- Aussagekräftig ist die relative Verteilung vor allem auch dann, wenn die Lage der Verteilungen angeglichen wird.
- Dekomposition:

$$\frac{f(y_r)}{f_0(y_r)} = \frac{f_A(y_r)}{f_0(y_r)} \times \frac{f(y_r)}{f_A(y_r)}$$

Overall relative density = Density ratio for the location difference  $\times$  Density ratio for the shape difference

z.B. mit  $F_A(y) = F_0(y + \rho)$

und  $\rho = \text{median}(Y) - \text{median}(Y_0)$



## Einige Masszahlen

- Gesamt-Unterschied der Verteilungen. Z.B. Kullback-Leibler Divergenz (als Informationsmass für die Abweichung der relativen Dichte von der Gleichverteilung):

$$D(F; F_0) = \int_{-\infty}^{\infty} \log \left( \frac{f(x)}{f_0(x)} \right) dF(x) = \int_0^1 \log(g(r)) g(r) dr.$$

- Polarisierung:

$$MRP(F; F_0) = 4 \int_0^1 \left| r - \frac{1}{2} \right| g_0^A(x) dr - 1.$$

$$LRP(F; F_0) = 8 \int_0^{1/2} \left| r - \frac{1}{2} \right| g_m(x) dr - 1$$

$$URP(F; F_0) = 8 \int_{1/2}^1 \left| r - \frac{1}{2} \right| g_m(r) dr - 1,$$

Summary Statistics for the Location/Shape Decomposition of the Relative  
Distribution of Wage Gains: Recent to Original NLS Cohort

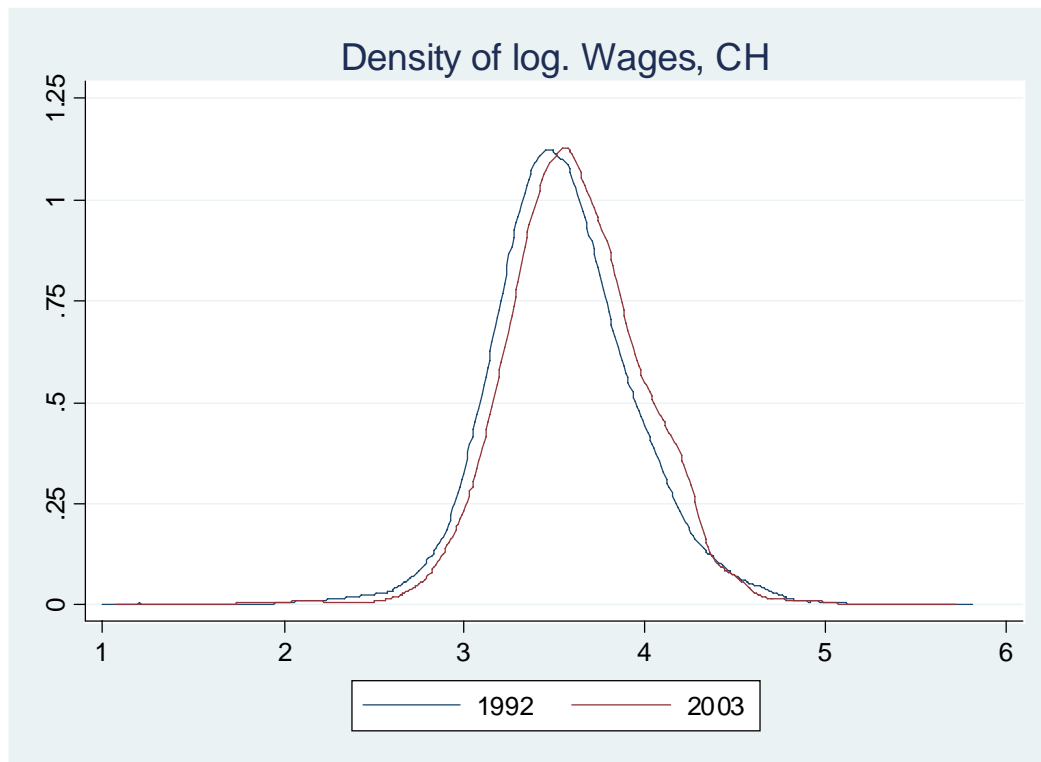
Entropy	Estimate	95% CI	<i>p</i> -value
Overall change in wage growth	0.125	0.092–0.158	0.000
Median effect	0.078	0.044–0.111	0.000
Shape effect	0.047	0.028–0.066	0.000
Percent due to median	62.4%	47.3–77.5	0.000
Percent due to shape	37.6	22.5–52.7	0.000
Polarization index	Estimate	95% CI	<i>p</i> -value
Median index	0.183	0.148–0.218	0.000
Lower index	0.190	0.111–0.270	0.000
Upper index	0.176	0.097–0.256	0.000

Quelle: Handcock & Morris (1998:74)

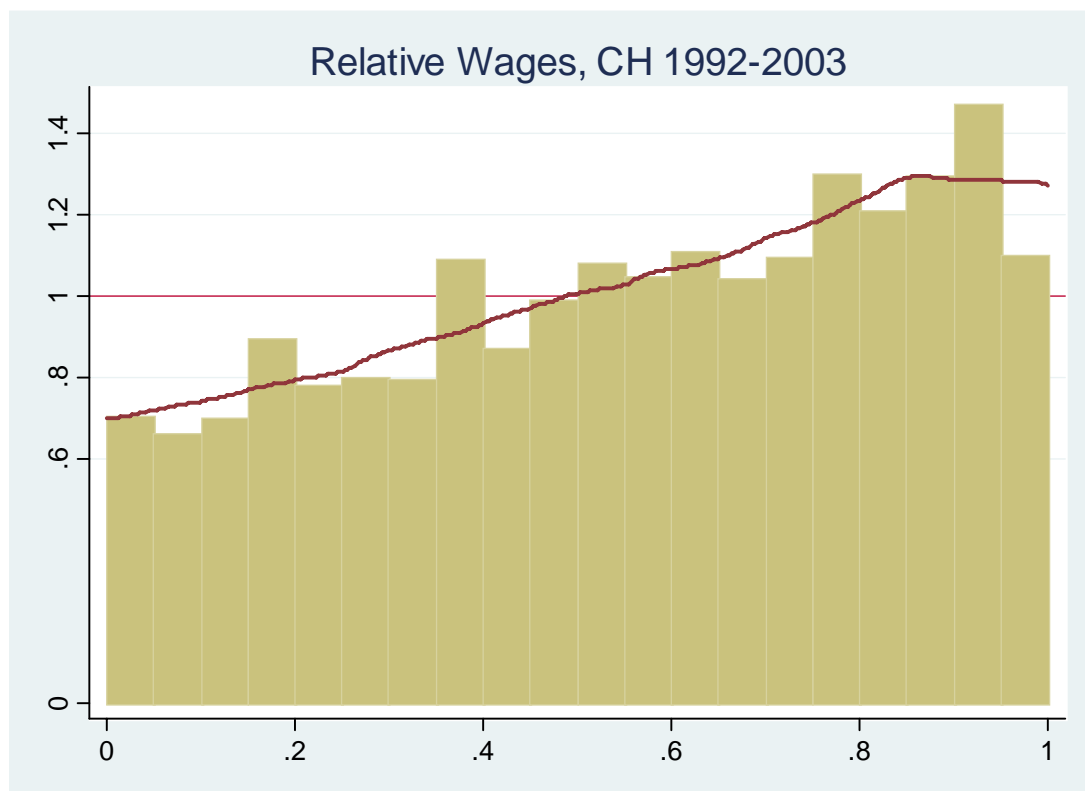
## Weitere Aspekte

- Kontrolle von Drittvariablen:
  - kontrafaktische Verteilungen können durch Gewichtung mit der relativen Dichte von X simuliert werden.
  - verwandter Ansatz: Matching
  - Allg. Problem: Effekte von einzelnen Kovariaten müssen sequenziell ermittelt werden (Pfadabhängigkeit)
- Schätzung:
  - Benötigt werden gute Schätzer für die Dichte, die zudem Berücksichtigen, dass die relativen Daten auf  $[0,1]$  begrenzt sind. Resultate können durch unterschiedlich starke Glättung stark beeinflusst werden. Ansätze: Histogramm-Schätzer, Kerndichte-Schätzer mit Randkorrektur, mixed-effects Poisson-Regression, polynomial Splines, ...
  - Inferenzstatistik z.T. ungelöst. Verwendung von Replikations-Methoden (bootstrap, jackknife).

# Einkommensentwicklung in der Schweiz 1992-2003

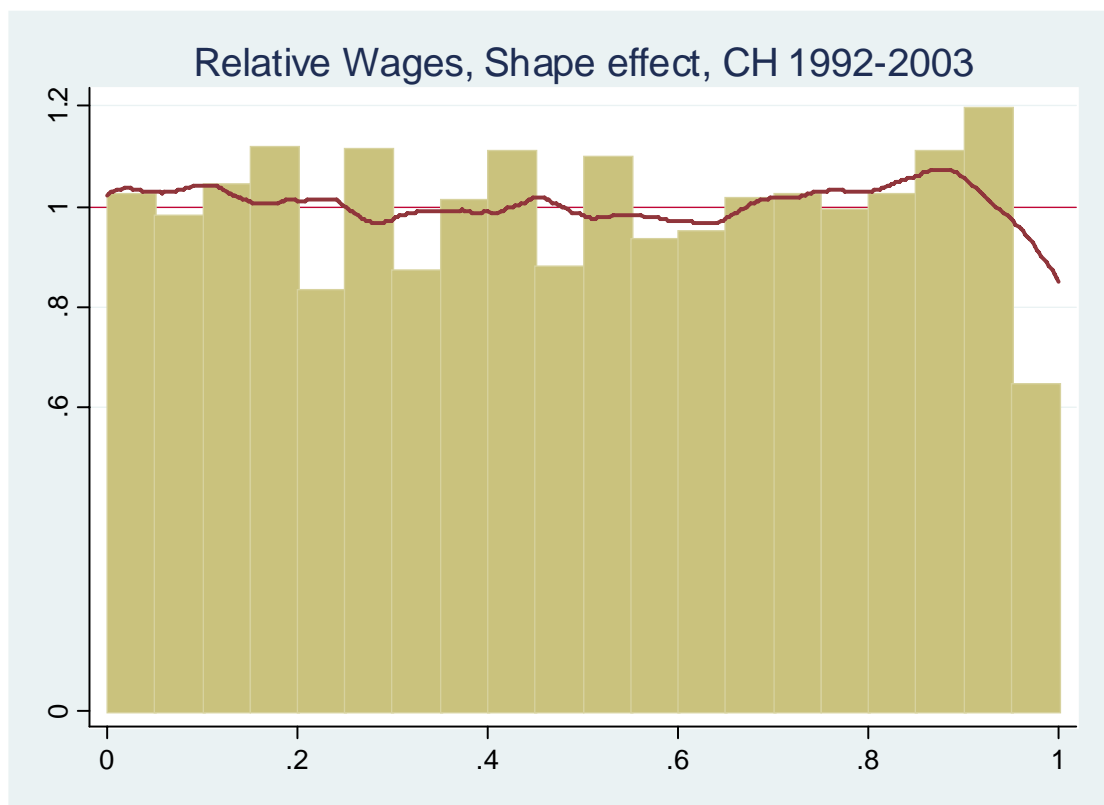


Quelle: Schweizerische Arbeitskräfteerhebung (SAKE)

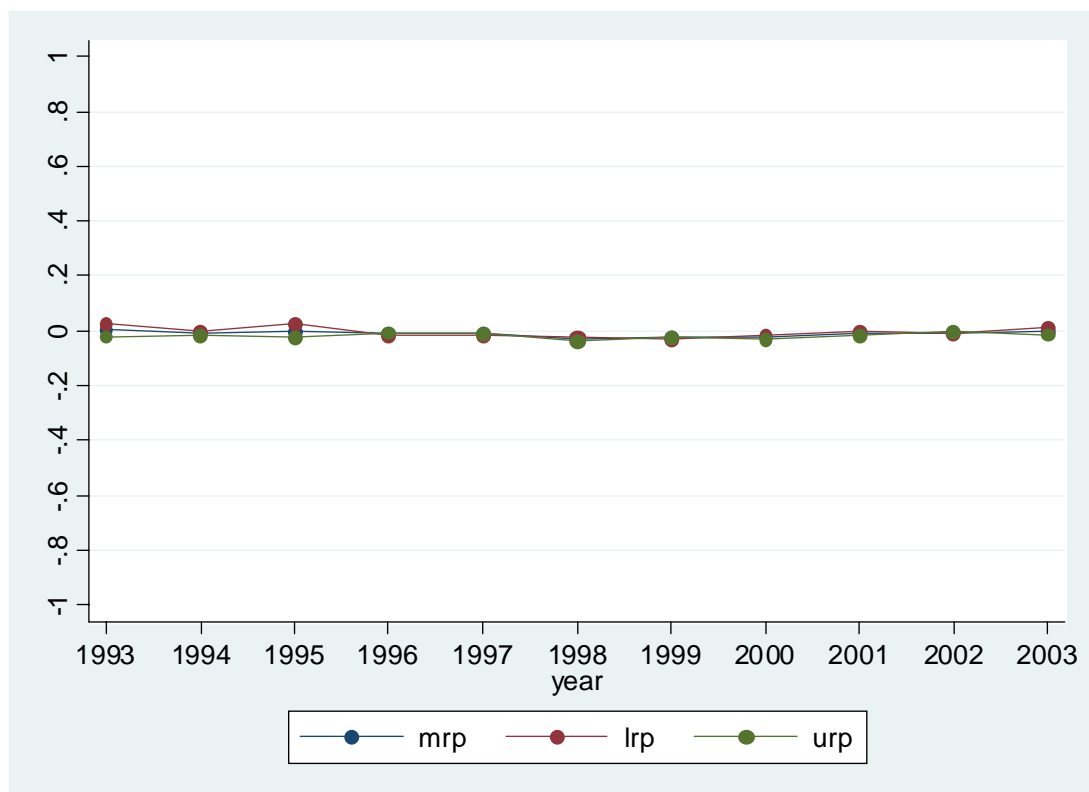


Quelle: Schweizerische Arbeitskräfteerhebung (SAKE)

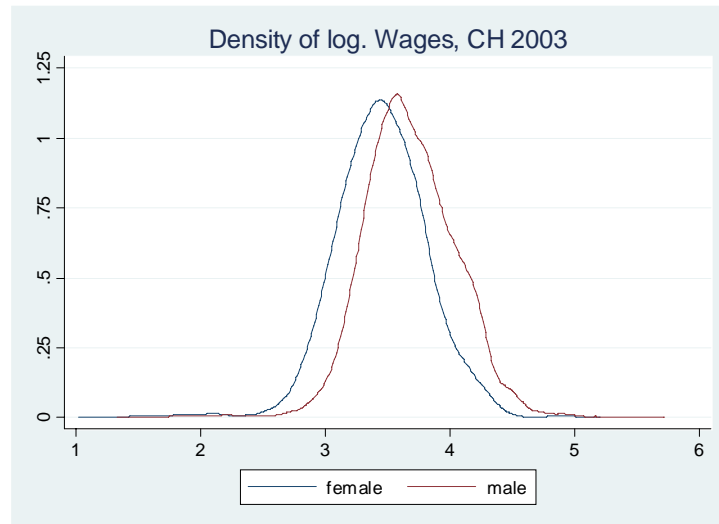
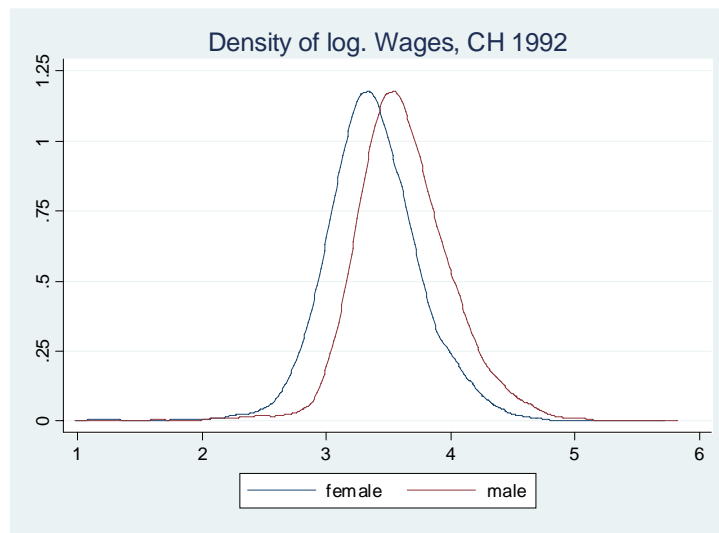




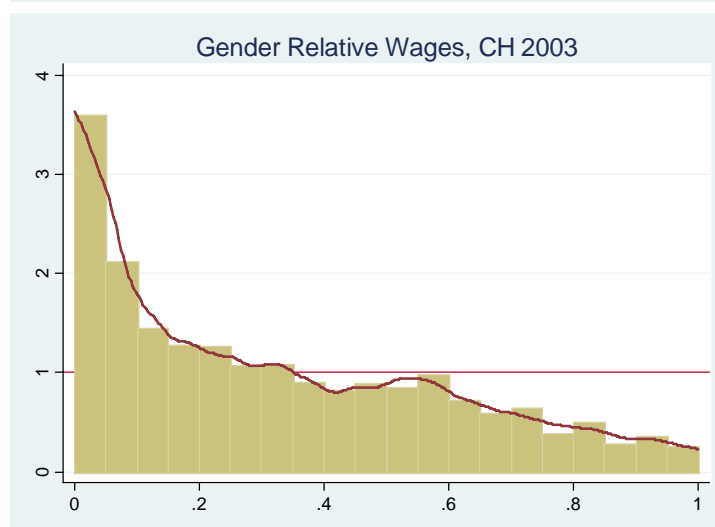
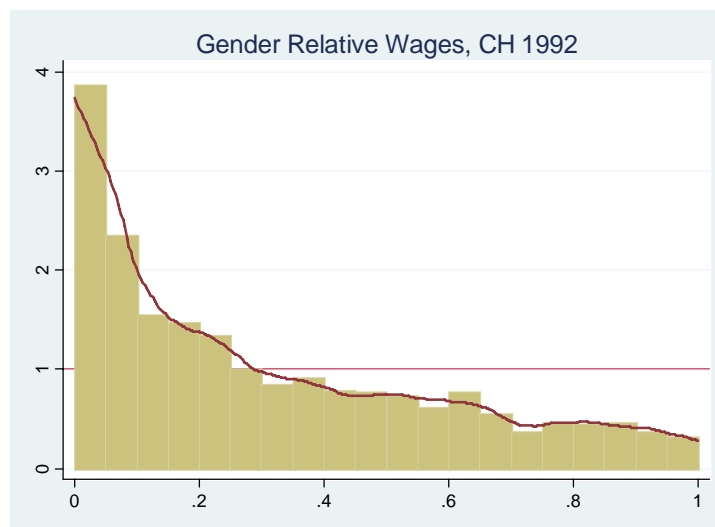
Quelle: Schweizerische Arbeitskräfteerhebung (SAKE)



Quelle: Schweizerische Arbeitskräfteerhebung (SAKE)

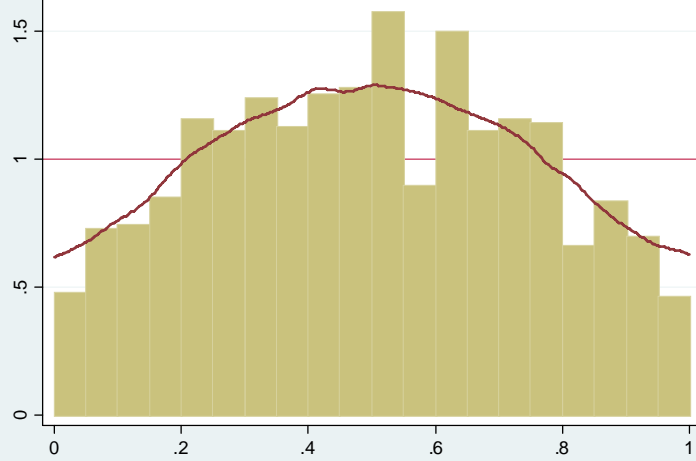


Quelle: Schweizerische  
Arbeitskräfteerhebung (SAKE)

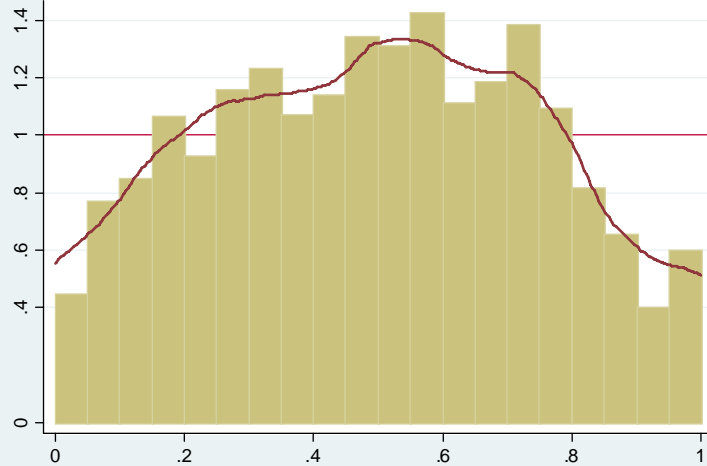


Quelle: Schweizerische  
Arbeitskräfteerhebung (SAKE)

Gender Relative Wages, Shape Effect (additive), CH 1992

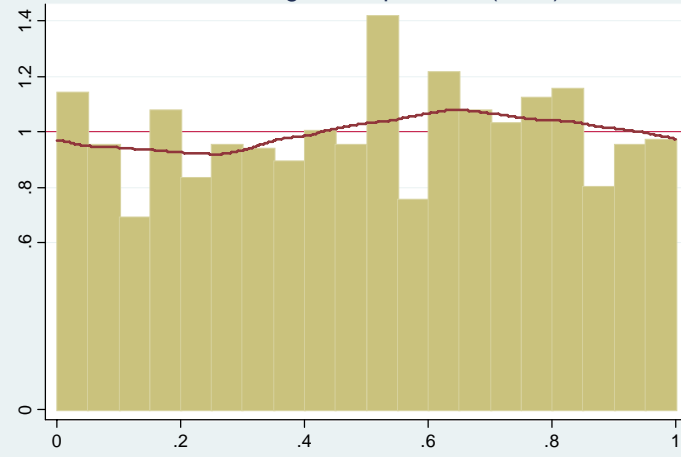


Gender Relative Wages, Shape Effect (additive), CH 2003

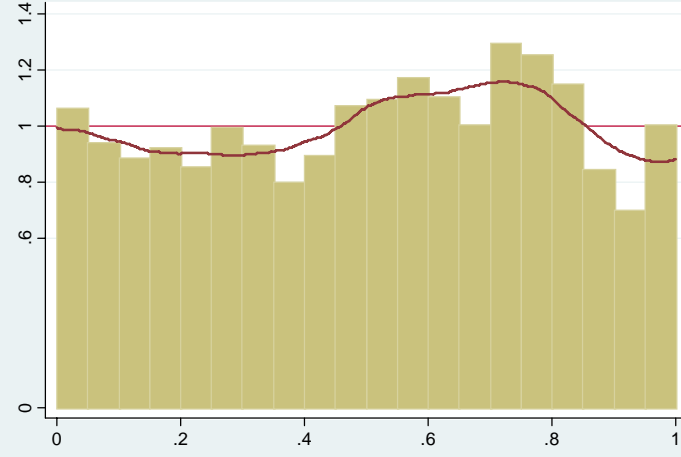


Quelle: Schweizerische  
Arbeitskräfteerhebung (SAKE)

Gender Relative Wages, Shape Effect (mult.), CH 1992



Gender Relative Wages, Shape Effect (mult.), CH 2003



Quelle: Schweizerische  
Arbeitskräfteerhebung (SAKE)

# Vielen Dank für Ihre Aufmerksamkeit!

## Literatur

- Bernhardt, A., M. Morris, and M. S. Handcock (1995). Women's Gains or Men's Losses? A Closer Look at the Shrinking Gender Gap in Earnings. *American Journal of Sociology* 101(2): 302-328.
- Blau, F. D., and Lawrence M. Kahn (1992). The Gender Earnings Gap: Learning from International Comparisons. *American Economic Review* 82(2): 533-538.
- Blau, F. D., and L. M. Kahn (1996). International Differences in Male Wage Inequality: Institutions versus Market Forces. *Journal of Political Economy* 104(4): 791-837.
- Blau, F. D., and L. M. Kahn (1996). Wage Structure and Gender Earnings Differentials: an International Comparison. *Economica* 63(250): S29-S62.
- Blau, F. D., and L. M. Kahn (1997). Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. *Journal of Labor Economics* 15(1): 1-42.
- Blinder, Alan S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources* 8(4): 436-455.
- DiNardo, J. E., N. Fortin, and T. Lemieux (1996). Labour Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica* 64(5): 1001-1046.
- Handcock, M. S., and M. Morris (1998). Relative Distribution Methods. *Sociological Methodology* 28: 53-97.
- Handcock, M. S., and M. Morris (1999). *Relative Distribution Methods in the Social Sciences*. New York: Springer.
- Juhn, C., K. M. Murphy, and B. Pierce (1991). Accounting for the Slowdown in Black-White Wage Convergence. Marvin Kosters (Hg.). *Workers and Their Wages*. Washington, DC: AEI Press.
- Juhn, C., K. M. Murphy, and B. Pierce (1993). Wage Inequality and the Rise in Returns to Skill. *Journal of Political Economy* 101(3): 410-442.
- Machado, J. A. F., and J. Mata (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics* 20(4): 445-465.
- Morris, M., A. D. Bernhardt, and M. S. Handcock (1994). Economic Inequality: New Methods for New Trends. *American Sociological Review* 59(2): 205-219.
- Nopo, H. (2004). Matching as a Tool to Decompose Wage Gaps. IZA Discussion Paper No. 981.
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review* 14(3): 693-709.